الاحصاء الاستدلالي INFERENTIAC STATIATICS

HYPOTHESIS TESTING اختبار الفروض

*** step of hypothesis testing** **لاختبار الفروض هناك 5 خطوات**

(1) state the null (H_0) and alternative (H_A) hypotheses

(1) وضع فرضية العدم (null hypothesis) H_0 وعادة ما يكون به علامة = (عدم وجود فرق)
(يفترض إنه لا يوجد اختلاف أو عدم وجود علاقة لأ يوجد تأثير .....)

وفرض البديل (alternative hypothesis)(H_A) وعادة لا يحتوي على علامة = ، وهو الفرض الذي يجب أن يكون صحيحا اذا كان فرض العدم غير صحيح (أ ي يقبل فرض البديل إذا رفض فرض العدم) .

فرض البديل (H_A) هو ما يود الباحث ان يثبت صحته ، وفرض العدم(H_0)هو ما يود الباحث ان يثبت ضدة

	Null Hypothesis H_0	Alternate Hypothesis H_A
Two – tailed	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$
One – tailed	$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$
	$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$

في حالة نكر بالسؤال هل النساء F أكثر وزنا من الرجال M يكونone – tailed ويكون الفرض

$H_0 : F \leq M$ $F \geqslant M$

$H_A : F > M$ $F < M$

نفس السؤال إذا كان بالسؤال هل يوجد اختلاف(زيادة .. نقص ..... تأثير) فيكون two – tailed ويكون الفرض

$H_0 : F = M$

$H_A : F \neq M$

tow tailed

يكون on tail إذا كان في السؤال أفعل تفضيل مثل أفضل ، أصغر ، أقل ، منخفض ، ..... lower – less – smaller –better)

$H_0 : F \leq M$ $H_A : F > M$

$H_0 : F \geq M$ $H_A : F < M$

**Women better than men

ويكون tow tail إذا كان هناك كلمة difference - affect (يؤثر - مختلف)

$H_0 : F = M$ $H_A : F \neq M$

The following describes a two-tailed hypothesis:

a) $A \geq B$.
b) $A \leq B$.
c) $A > B$.
d) $A = B$..
e) $A < B$.

(2) وضع قيمة α وعادة تكون 0.05 وهو الخطأ المحتمل (2) set the significance level(α)
أي ان الخطأ 0.05 يعني نقبل الخطاء بنسبة 0.05 لكي نقبل النتيجة بنسبة 95%

(3) اختيار الاختبار الإحصائي المناسب

(3) select the appropriate test of significance , calculate the test statistic.

الاختبار الاحصائي يعتمد على نوع المتغيرات فإذا كان :

1 الاختبار الأول نوعى (Qualitative) والثاني نوعي (Qualitative) نستخدم اختبار

Chi square (X^2) test المقارنة بين الرجال والنساء ، الأولاد والبنات .....

2 إذا كان الأول كمي (Quantitative) والثاني نوعي (Qualitative) نستخدم اختبار **t – test**

وهناك نوعان لـ **t- test** لتحديد تأثير التدخين على الوزن و

1) اختبار t المزدوج **paired t- test** واختبار t الغير مزدوج **Unpaired t- test**

إذا كان هناك مجموعة واحدة نستخدم اختبار **paired t- test** وإذا كان هناك مجموعتين نستخدم اختبار **Unpaired t- test**

2) إذا كان المتغيرين كلاهما كمي Quantitative نختار اختبار الارتباط Correlation

(4) مقارنة القيمة المحسوبة بالقيمة الجدولية

(4) compare the Probabiiitv value to significance level

(2) calculated value< tabular value **(accept the null hypothesis)**

1- إذا كانت القيمة المحسوبة أقل من القيمة الجدولية نقبل فرض العدم ويكون الاختبار ذو دلالة إحصائية

then **accept the H_0** and conclude that is Statistically significant.

(1) calculated value > tabular value **(Reject the null hypothesis)**

2- إذا كانت القيمة المحسوبة أكبر من القيمة الجدولية نرفض فرض العدم ونقبل الفرض البديل ويكون الاختبار ليس له دلالة إحصائية .

then **reject the H_0** and conclude that is **not** statistically significant.

---

reject نرفض **Fail to reject - accept** نقبل

(5) interpret results (conclusion) (5) الخلاصة (النتيجة)

اختبار كاي تربيع Chi- Square (X^2) test (1)

هو اختبار احصائي نطبقه لاختبار الفروض للمتغيرات النوعية مثل المقارنة بين الرجال والنساء من حيث....

The key idea of the Chi- square test is a comparison of **observed** and **expected** values . وتعتمد فكرة هذا الاختبار على المقارنة بين القيم المتوقعة (**expected frequency**) والقيم المشاهدة (**observed frequency**).

ويتم حساب كاي تربيع كما يلي

$$x^2 = \sum \frac{(\text{observed frequency} - \text{expectd frequency})^2}{\text{expectd frequency}}$$

القيمة المشاهدة — القيمة المتوقعة

$$\frac{(\text{القيمة المتوقعة} - \text{القيمة المشاهدة})^2}{\text{القيمة المتوقعة}}$$

To determine the significance level we need to " degree " freedom (df)

لتحديد مستوى المعنوية يجب حساب درجة الحرية $df = (n-1)(n-1)$ df

حيث n هي عدد الصفوف ، n الأعمدة (degree freedom)= $(\text{Row}-1)(\text{column}-1)$

ملاحظات

في اختبار كاي يكون (فرض العدم H_0) حيث $H_0: \mu_1 = \mu_2$

و (فرض البديل H_A) $H_A: \mu_1 \neq \mu_2$

نفرض $\alpha = 0.05$

نحسب قيمة X^2 التي تتطلب حساب القيم المتوقعة والقيم المشاهدة بعد ذلك نحسب درجة الحرية وأخيرا نقرر ما إذا كان نقبل الفرض أو نرفض الفرض .

Example (1)

The results of a biostatistics exam showed that out of 120 male students, 20 had "F", while out of 80 female students 10 had "F". Are the results of male students **different** from those of female students? (=) Two-wailed

أظهرت نتائج امتحان الإحصاء الحيوي أن من أصل 120 طالبا، كان 30 "F"، في حين من أصل 80 طالبة كان10 "F". هل نتائج الطلاب الذكور تختلف عن تلك الطالبات؟

Observed frequency القيم المشاهدة

	F	A	Total
Mal	20	100	120
Female	10	70	80
Total	30	170	200

$H_0: \mu_1 = \mu_2$

$H_A: \mu_1 \neq \mu_2$

فرض العدم ينص على ان التساوي النتائج بين البنين والبنات

من هذا الجدول يمكن حساب القيمة المتوقعة لكل خلية في الجدول وذلك بضرب

مجموع الصف × مجموع العمود وقسمة الناتج على المجموع الكلي $\frac{\text{مجموع الصف} \times \text{مجموع العمود}}{\text{المجموع الكلي}}$

الطلاب الذكور الذين حصلوا على معدل F مثلا $\frac{120 \times 30}{200} = 18$

بالمثل الطالبات البنات اللاتي حصلن على F

expected frequency القيم المتوقعة

	F	A	Total
Mal	18	102	120
Female	12	68	80
Total	30	170	200

$df = (2-1)(2-1) = 1 \times 1 = 1$ degrees " freedom (d f) درجة الحرية

حيث 2 هي عدد الأعمدة و 2 هي عدد الصفوف

ثم من جدول Chi- Square نختار الصف الذي درجة الحرية 1 تحت العمود 0.05 نجد القيمة

3.841 القيمة الجدولية

ثم نوجد القيمة المحسوبة(calculated value) بالقانون

$$\frac{(\text{القيمةالمتوقعة} - \text{القيمةالمشاهدة})^2}{\text{القيمة المتوقعة}}$$

$$x^2 = \frac{(20-18)^2}{18} + \frac{(100-102)^2}{102} + \frac{(70-68)^2}{68} + \frac{(10-12)^2}{12} = \frac{100}{153} = 0.653$$

$$= \frac{4}{18} + \frac{4}{102} + \frac{4}{68} + \frac{4}{12} = 0.653$$ الحل

ثم نقارن القيمة الجدولية (3.841) بالقيمة المحسوبة (0.653) نجد أن

نجد أن القيمة المحسوبة اقل من القيمة الجدولية وبذلك نقبل فرض العدم

accept the null hypothesis

ملحوظة إذا كانت القيمة المحسوبة (calculated value) اقل من القيمة الجدولية (tabular value)

accept the null hypothesis تقبل فرض العدم ونرفض الفرض البديل

إذا كانت القيمة المحسوبة (calculated value) اكبر من القيمة الجدولية(tabular value)

Reject the null hypothesis نرفض فرض العدم ونقبل البديل

---

1) How many variables do we have? كم متغير لدينا

a) 5 .
b) 3 .
c) 1 .
d) 2.
e) 4 .

---

2) The test statistic to be applied is: الاختبار الإحصائي الذي يتعين تطبيقه هو

a) Unpaired t-test .
b) Chi Square test.
c) Correlation.
d) Regression .
e) Paired t-test .

3)The tabular value is:: القيمة الجدولية هي

a) 7.879.
b) 2.706 .
c) 3.841.
d) 6.635
e) 5.412.

---

4) The calculated value is:: القيمة المحسوبة هي

a) 1.9
b) 5.9 .
c) 0.565.
d) 2.9 .
e) 0.653 ..

---

5) The following is correct:: ما يلي هو الصحيح

a) Reject the H_0. نرفض فرض العدم
b) Accept H_A. نقبل فرض البديل
c) Accept H_0.. نقبل فرض العدم
d) The calculated value is higher than the critical (tabular) value. المحسوبة اكبر من الجدولية
e) The tabular(critical) value is equal to the calculated value. الجدولية تساوي المحسوبة

اختبار تا (غير المزدوج) (مجموعتين) Unpaired t- test (2)

كمي ونوعي

ملحوظة : إذا ذكر في السؤال – lower – less – smaller أو difference - affect يكون

الاختبار **t - test**

يعتمد اختبار (تا) على :

1. المتوسط الحسابي لكل مجموعة (mean)
2. التباين في كل مجوعة (variance)
3. الأعداد في كل مجموعة (number)

ويتم حساب قيمة t بالمعادلة

$$t = \frac{\overline{X_A} - \overline{X_B}}{\sqrt{\frac{Var_A}{nA} + \frac{Var_B}{nB}}}$$

حيث $\overline{X_A}$ هو المتوسط الحسابي للمجموعة الكبيرة و $\overline{X_B}$ هو المتوسط الحسابي للمجموعة الثاني

و Var_A هو التباين للمجموعة الأولى و Var_B هو التباين للمجموعة الثاني

ويمكن حساب درجة الحرية degree of freedom في اختبار t كما يلي

$$d f = n_1 + n_2 - 2 \longrightarrow d f = (n_1 - 1) + (n_2 - 1)$$

حيث n_1 هي عدد الحالات للمجموعة الأولى و n_2 هي عدد الحالات للمجموعة الثانية

وبعد الحساب إذا كانت :

القيمة المحسوبة < القيمة الجدولية ((نقبل فرض العدم **accept the null hypothesis**

القيمة المحسوبة > القيمة الجدولية (((نرفض فرض العدم **Reject the null hypothesis**

Example

A group of 4 persons (treatment group) claimed that they smoke to lower their body weight and avoid obesity . their body weights were compared with a group of 4 non- smoking can be applied to **lower** body weight?

هناك 4 شخص من المدخنين نصحوا بعدم التدخين فكان ردهم بأنهم يدخنون لتجنب السمنة وعند دراسة أوزان 4 شخص من غير المدخنين وأوزان 4 شخص من غير المدخنين وجدا التالي

المطلوب تحديد السؤال البحثي هل التدخين **يخفض** السمنة

ID	1	2	3	4
مدخنين Treatment Group	68	66	63	67
غير مدخنين Control Group	67	71	65	69

$df = (n_1 - 1) + (n_2 - 1)$

$= (4-1) + (4-1)$

$= 3 + 3 = 6$

$$\bar{X} = \frac{68+66+63+67}{4} = \frac{264}{4} = 66 \quad ,\bar{y} = \frac{67+71+65+96}{4} = \frac{272}{4} = 68$$

لحساب **variance**

$$v = \frac{\sum (x-\bar{x})^2}{n-1}$$

X	$x-\bar{X}$	$(x-\bar{x})^2$
68	2	4
66	0	0
63	-3	9
67	1	1
المجموع	0	14

y	$y-\bar{y}$	$(y-\bar{y})^2$
67	-1	1
71	3	9
65	-3	9
69	1	1
المجموع	0	20

$$v_x = \frac{\sum(X-\bar{X})^2}{n-1} = \frac{14}{4-1} = \frac{14}{3} = 4.67$$ تباين المجموعة الاول

$$v_y = \frac{\sum(y-\bar{y})^2}{n-1} = \frac{20}{4-1} = \frac{20}{3} = 6.67$$ تباين الثانية

الكبير — الصغير

$$t = \frac{\overline{X_A} - \overline{X_B}}{\sqrt{\frac{Var_A}{nA} + \frac{Var_B}{nB}}} \qquad t = \frac{68-66}{\sqrt{\frac{4.67}{4} + \frac{6.67}{4}}} = \frac{2}{\sqrt{1.17+1.67}} = \frac{2}{\sqrt{2.84}} = 1.186$$

Mean الوسط الحسابي	$\frac{264}{4} = 66$	$\frac{272}{4} = 68$
Variance التباين	$\frac{14}{3} = 4.67$	$\frac{20}{3} = 6.67$
$(\overline{X_A} - \overline{X_B})$	الكبير ناقص الصغير 68 - 66 = 2	
Calculated t - test القيمة المحسوبة	1.186 من القانون	
d f درجة الحرية	$n_1 + n_2 - 2$ 4 + 4 - 2 =6	
Type of test نوع الاختبار	One – tailed	
Tabular t- value القيمة الجدولية	1.943 من الجدول	
Decision القرار	**accept the null hypothesis**	

القيمة المحسوبة اقل من القيمة الجدولية لذلك نقبل الفرض **accept the null hypothesis** الذي ينص على إن وزن المدخنين أقل من غير المدخنين .

---

4) The calculated value is:: القيمة المحسوبة هي

a) 1.434
b) 2.168
c) 1.186..
d) 2.934 .
e) 0.656 .

---

5) The following is correct:: ما يلي هو الصحيح

a) Reject the H_0. نرفض فرض العدم
b) Accept H_A.
c) accept the null hypothesis قبول فرض العدم
d) The calculated value is higher than the critical (tabular) value. المحسوبة اكبر من الجدولية
e) The tabular(critical) value is equal to the calculated value. الجدولية تساوي المحسوبة

paired t- test (3) اختبار(تا) (المزدوج) (مجموعة واحدة)

يختلف اختبار t المزدوج عن اختبار t غير المزدوج في انه يشمل مجموعة واحدة .

أي أن يكون هناك مجموعة واحدة وكل شخص يكون له قراءتين الأولى قبل العلاج والثانية بعد العلاج أو مقارنة الجزء الأيمن بالجزء الأيسر.

مثال لذلك ضغط الدم **قبل** العلاج **وبعد** العلاج لمجموعتين أو تأثير العلاج لمريض السكر قبل العلاج وبعدة

مقارنة تأثير عقار معين ، مقارنة وزن اليد اليمنى عن وزن اليد اليسرى لمجموعة أشخاص أو حدة الأبصار للعين اليمنى عنها للعين اليسرى .

ونستخدم القانون $$t = \frac{\overline{d}}{\sqrt{\frac{s2}{n}}}$$

حيث **d** هي الفرق بين القراءتين $\overline{d}$, المتوسط الحسابي

degree of freedom = n - 1

Example

The weights of 8 children were measured before and after treatment with a certain drug are presented in following table

تم قياس أوزان 8 أطفال قبل وبعد العلاج بدواء معين وكانت القراءات كما في الجدول التالي

Child : ID	1	2	3	4	5	6	7	8
Before treatment	32	16	37	32	44	28	27	48
After treatment	38	10	30	36	50	20	18	38

Tw0 – tail الاختبار هل الوزن زاد حقا بعد العلاج.

Test if the weight of these 8 children has really **increased** after treatment .

	d	$d-\bar{d}$	$(d-\bar{d})^2$
	-6	-6-3=-9	81
	6	6-3=3	9
	7	4	16
	-4	-7	49
	-6	-9	81
	8	5	25
	9	6	36
	10	7	49
Total	24	0	346

$d = B-A = 32-38 = -6$
$= 16-10 = 6$
$= 37-30 = 7$
$= 32-36 = -4$
$= 44-50 = -6$
$= 28-20 = 8$
$27-18 = 9$
$-48-48 = 10$
24

$$\bar{d} = \frac{-6+6+7-4-6+8+9+10}{8} = \frac{24}{8} = 3$$

لحساب التباين

$$s^2 = \frac{\sum(d-\bar{d})^2}{n-1} = \frac{346}{8-1} = 49.42$$ التباين

$$t = \frac{\bar{d}}{\sqrt{\frac{s2}{n}}} = \frac{3}{\sqrt{\frac{49.42}{8}}} = 1.207$$ القيمة المحسوبة

6.1775

$df = n-1 = 8-1 = 7$, $\alpha = 0.05$

Tabular value = **2.364** الجدولية

accept the null hypothesis القيمة المحسوبة اقل من الجدولية لذا نقبل فرض العدم

1) **The suitable statistical test to answer this problem is :**
   a) Chi square.
   b) Unpaired t-test.
   c) Regression .
   d) Paired t-test..
   e) Correlation

---

2) **The degree of freedom.**
   a) 5.
   b) 6
   c) 7..
   d) 8 .
   e) 14.

المحسوبة

3) **The calculated test-value is :**
   a) 2.201 .
   b) 1.980.
   c) 2.457.
   d) 1.207..
   e) 2.365.

---

4) **your conclusion is:** الاستنتاج
   a) accept the null hypothesis. H_0
   b) Accept H_A.
   c) Reject the H_0.
   d) The calculated value is higher than the critical (tabular) value
   e) The tabular(critical) value is equal to the calculated value.

إذا كانت القيمة المحسوبة أقل من الجدولية نقبل

أكبر نرفض

(4) CORRELATION(4) الارتباط

عند قياس متغيرين كميين ، فهل المتغيرات تتغير مع بعض أم لا .

إذا ما تغير الأول فهل يتغير الثاني تبعاً لما يتغيره الأول ؟

وإذا كان هناك تغير مشترك بينهما فما مقدار هذا الاشتراك في التغير ؟

** الارتباط هو إجراء لتحديد درجة الاتحاد ودرجة الإشراك لمتغيرين وقياس هذا الاتحاد بمقياس يتراوح بين +1 to 0 through -1.

ويشترط لمقياس الارتباط أن يكونا المتغيرين مستقلين عن بعضهما البعض ويكونا المتغيرين موزعين توزيعاً طبيعياً .

ويرمز لمعامل الارتباط بالرمز (**r**) وتتراوح قيمته بين +1 to -1

- عندما يكون **(+1)** فهذا يعني ارتباط تام موجب perfect liner positive correlation
- عندما يكون **(-1)** فهذا يعني ارتباط تام سالب perfect linear negative correlation
- عندما يكون **(0)** فهذا يعني لا يوجد ارتباط No correlation
- إذا كانت r قريبة من الواحد (0.7, 0.8 , 0.9) فهذا يعني ارتباط قوي strong correlation
- إذا كانت r (0.1 ,0.2 , 0.3) فهذا يعني ارتباط ضعيف weak correlation
- إذا كانت r (0.4 , 0.5 , 0.6) فهذا يعني ارتباط متوسط median correlation

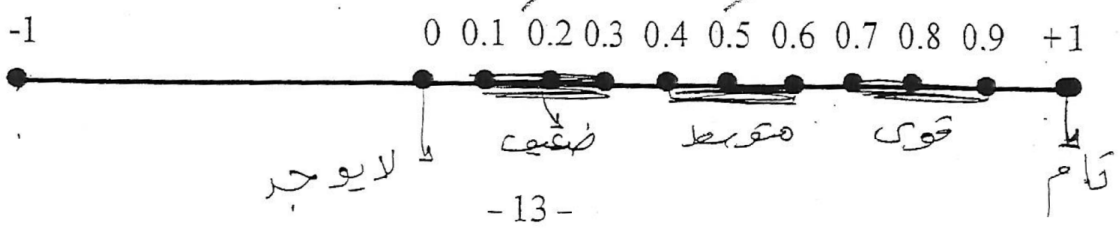
ويتم حساب الارتباط من المعادلة

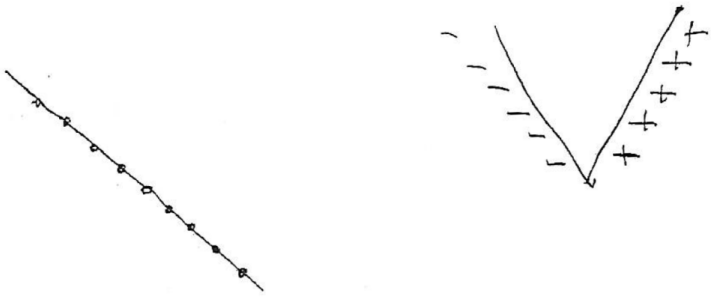
$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

أمثلة على الارتباط :

العمر وساعات النوم للطفل (علاقة عكسية) الاشارة سالبة

العمر وطول الطفل (علاقة طردية) الاشارة موجبة





Perfect negative (r = -1)

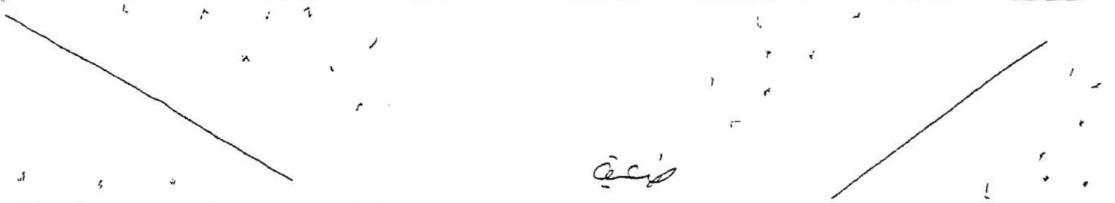
تام

Perfect positive(r= 1)

Strong negative(r =- 0.7 , -0.8 , -0.9)

قوي

strong positive(r = 0.7 , 0.8 , 0.9)



Weak negative(r =- 0.1 ,- 0.2 , -0.3)

ضعيف

Weak positive(r = 0.1 , 0.2 , 0.3)

Median negation(r = -0.4 , -0.5 , -0.6)

median positive(r= 0.4 , 0.5 , 0.6)

لا يوجد

No correlation (r = 0)

Example

The following table shows the data of height (x) and weight (y) of 5 children " r" between both of these two variables

يبين الجدول التالي البيانات بين ارتفاع ووزن 5 أطفال . والمطلوب إيجاد الارتباط بين المتغيرين

height	20	20	26	18	16
weight	105	110	120	90	75

height x	**Weight** y	$x-\bar{X}$	$y-\bar{Y}$	$(x-\bar{X})(y-\bar{Y})$	$(x-\bar{x})^2$	$(y-\bar{Y})^2$
20	105	0	5	0	0	25
20	110	0	10	0	0	100
26	120	6	20	120	36	400
18	90	-2	-10	20	4	100
16	75	-4	-25	100	16	625
$\bar{X}=20$	$\bar{y}=100$	0	0	240	56	1250

$$\bar{X} = \frac{20+20+26+18+16}{5} = \frac{100}{5} = 20$$

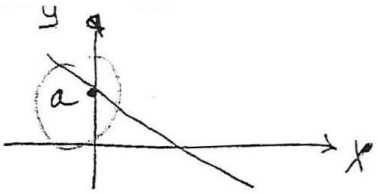
$$\bar{y} = \frac{105+110+120+90+75}{5} = \frac{500}{5} = 100$$

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{Y})^2}}$$

$$r = \frac{240}{\sqrt{56 \times 1250}} = 0.907$$

strong positive correlation فهذا يعني ارتباط قوي

(5) Regression الانحدار



Linear regression is a statistical procedure for predicting the value of a dependent variable from an independent variable when the relationship between then.

الانحدار الخطي هو إجراء احصائي يمكننا من التنبؤ بقيمة المتغير التابع بمعلومية المتغير المستقل عندما تكون العلاقة بينهما علاقة خطية .

Regression is fitting of the best line through a series of points .

الانحدار الخطي هو رسم خط أفضل ما يمكن خلال مجموعة من النقاط .

يمكن تمثيل معادلة الانحدار الخطي بالصورة $\hat{y} = a + b\,x$

x هو قيمة المتغير المستقل independent variable .

a هو قيمة intercept نقطة تلاقي خط الانحدار مع المحور الرأسي .

b هو الميل slop ، (a ,b) ثوابت (constants)

y هو المتغير التابع (dependent)

Example :

During infancy (first 12months of life) , the length of the baby can be predicted according to the following regression equation:

Length of the baby (1n cm) = 50 + (2 x age in months)

Can you predict the length of the baby at age of 7 months?

في مرحلة الطفولة في اول 12 شهر في الحياة يمكن توقع طول الرضيع وفقاً لمعادلة الانحدار التالية
هل يمكن التنبؤ بطول الرضيع بعد 7 شهور ؟ -

1) The Length of the baby (in cm) at age 7 month .

a) 46 cm .
b) 52 cm .
c) 64 cm ..
d) 66 cm .
e) 36 cm .

2) **The Length of the baby (in cm) at age 13 month .**

a) 76 cm .
b) 77 cm .
c) 75 cm .
d) 74 cm .
e) None of the above is correct..

لا يصلح أنه نعوض بـ 13 لأنه أكبر من 12 المعادلة صالحة لغاية 12 شهر .

---

3) **We consider the age in months** نعتبر العمر بالشهر

a) dependent variables.
b) slop .
c) Independent variables..
d) intercept .
e) Qualitative .

---

4) **The study the relationship between " height " and " age " in children, we would use :**

a) Chi square.
b) Unpaired t-test.
c) Regression .
d) Paired t-test.
e) Correlation..

---

1) **What is the predicted systolic blood pressure (mmHg) for a women 45 Kg Where a = 79.7 and b = 0.8**

a) 125 mmHg .
b) 115 mmHg .
c) 100 mmHg .
d) 111.7 mmHg .
e) 115.7mmHg ..

=======*******=======*******=======

=======*****=======*****=======

(Mohamed ezzat) مع اطيب الأمنيات بالنجاح